2Д

УДК 004.853

Д.Г. Досин, Р.Р. Даревич, Н.В. Шкутяк
Фізико-механічний інститут імені Г.В. Карпенка НАН України, м. Львів, Україна
dosyn@ipm.lviv.ua, darevych@ipm.lviv.ua, nyk_nata@ipm.lviv.ua

Розробка онтології матеріалознавства засобами Protégé-OWL

У статті на основі існуючих підходів до побудови онтологій нижнього рівня, а також форматів та стандартів подання знань представлено розроблену авторами технологію побудови онтології матеріалознавства засобами Protégé-OWL. Розроблена онтологія орієнтована на застосування в інтелектуальних системах інформаційного пошуку у текстових масивах даних. Проаналізовано можливості практичного застосування розробленої онтології у наукових дослідженнях.

Вступ

Сьогодні надзвичайно актуальне питання створення інтелектуальних систем у певних галузях науки та промисловості, котрі могли б значною мірою спростити інженерам та науковцям пошук та доступ до потрібної (релевантної) інформації у сховищах величезних об'ємів. Систему, що вирізняється вмінням визначати, зберігати та використовувати в потрібний момент необхідні релевантні знання, називатимемо інтелектуальною системою (ІС). В основі архітектури сучасних ІС лежать онтології, що формуються для заданої предметної області (ПрО). У процесі реалізації ІС найскладнішим є етап початкового формування такої онтології та подальшого її наповнення новими знаннями.

Успіх у розв'язанні задачі побудови ефективної спеціалізованої ІС визначається відповідністю її бази знань та онтології (ядра бази знань) до особливостей ПрО. З огляду на це, в області інформаційних технологій зростає актуальність проблеми розробки таких онтологій, які можна застосовувати для роботи інтелектуальних пошукових машин в галузі інформаційного пошуку, та експертних систем, орієнтованих на певну ПрО [1], [2]. На сьогодні розроблено ряд онтологій, які призначені для використання у промисловості в складі інтелектуальних систем, функцією котрих є прийняття рішень під час вирішення конкретних завдань. Як приклад, можна навести побудову та використання онтології в складі системи керування знаннями в металопромисловості в Тайвані [3]. Цю онтологію використовують в системі керування знаннями на етапі маніпулювання і підтримки завдання пошуку та керування. Вона містить знання про методи та технології виплавляння металів і сплавів, однак охоплює лише певну частину галузі матеріалознавства – металургію, а також не забезпечує в достатній мірі відображення властивостей матеріалів. Ряд розроблених універсальних ІС [4], [5] базуються на онтології верхнього та середнього рівня і охоплюють одразу декілька ПрО, проте в наукових дослідженнях необхідні вузькоспеціалізовані прикладні системи, здатні містити найбільш повну інформацію про об'єкт досліджень.

Метою статті є розроблення онтології в галузі матеріалознавства для по треб науковців Фізико-механічного інституту ім. Г.В. Карпенка НАН України (http://www.ipm.lviv.ua), оскільки тут проводяться широкі дослідження в галузі міцності і довговічності матеріалів під впливом механічних навантажень, температур і корозійно-

активних середовищ, технологій захисту металів від корозії, методів і засобів технічної діагностики матеріалів. Така онтологія повинна охоплювати фундаментальні знання матеріалознавства як науки, а також нові знання, одержані в цій галузі.

Огляд існуючих засобів побудови онтології

На сьогодні розробка онтологій, як явних формальних описів понять ПрО та зв'язків між ними, переходить з лабораторій штучного інтелекту на робочі столи експертів з певних предметних областей − інженерів та користувачів. Це стало можливим завдяки використанню спеціалізованих редакторів, які дають змогу наповнювати онтологію новими знаннями безпосередньо самому спеціалісту з даної ПрО.

Для створення та редагування онтологій ІС розроблено ряд спеціалізованих середовищ розробки, редакторів, парсерів та засобів об'єднання онтологій, найбільш ефективними з яких є: KAON [6], OilEd [7], OntoEdit [8], Ontosaurus [9], OpenCyc [10], Protégé [11].

Серед цих інструментів для побудови предметно-орієнтованої онтології виділимо редактор Protégé-OWL [12], як гнучке, незалежне від платформи середовище з своїми особливостями та перевагами, яке забезпечує наочний та зручний у використанні графічний інтерфейс користувачу, реалізує масштабованість, тобто модульне нарощування системи в рамках уніфікованої архітектури, дає змогу нарощувати архітектуру за допомогою додатково розроблених підпрограм – плагінів (plug-in). Також Protégé-OWL дає змогу описувати класи з використанням нових можливостей. Зокрема, мова OWL (Ontology Web Language) має великий набір операторів і базується на логічній моделі, яка дозволяє давати визначення поняттям так, як вони описані, тому складні комплексні поняття у визначеннях можуть бути створені з простіших. До того ж логічна модель дає змогу використовувати механізм міркувань (Reasoner), котрий у свою чергу дає змогу перевірити чи твердження і визначення в онтології є взаємно несуперечливими, а також розпізнати відповідність визначень певним поняттям. Завдяки цьому механізму підтримується правильність ієрархії онтології.

Редактор також підтримує більшість реляційних баз даних (Oracle, MySQL, Microsoft SQL Server, Microsoft Access).

У редакторі Protégé-OWL забезпечено можливість вибору однієї з трьох розроблених на цей час версій мови OWL: OWL-Lite, OWL-DL, OWL-Full. Версії відрізняються мірою їх виразності. OWL-Lite є найпростішою з погляду синтаксису. Її найкраще використовувати під час побудови простої ієрархії класів та визначення семантичних обмежень. OWL-DL є значно виразнішою підмовою порівняно з OWL-Lite. Вона базується на описовій логіці (description logics) і орієнтована на тих користувачів, які хочуть максимальної прозорості без втрати повноти обчислень. OWL-DL охоплює всі мовні конструкції OWL з обмеженнями, на зразок розділення типів (клас не може бути властивістю, а властивість не може бути об'єктом або класом). Завдяки цьому OWL-DL забезпечує автоматичний логічний вивід (процес міркування). OWL-Full використовують в ситуаціях, де виразність засобів більш важлива, ніж потреба у обчислювальних можливостях мови. Вона забезпечує максимальну синтаксичну свободу стандарту RDF без обчислювальних гарантій. Водночас вона використовує всі переваги попередніх двох підмов.

У цілому OWL як мова web-онтологій забезпечує:
– синтаксис опису понять, зручний для всіх користувачів мережі Інтернет (людей та програмних агентів);
– максимальну виразність механізмів опису понять та зв'язків між ними;
– механізми еволюції описів та спільного використання онтологій в середовищі Інтернету.

Таким чином, для розроблення онтології в галузі матеріалознавства вибрано редактор Protégé-OWL. Нижче описано технологію створення онтології з використанням цього засобу, а також можливості її практичного застосування в складі інтелектуальної системи.

Технологія розробки онтології матеріалознавства

Для побудови онтології, котра реалізована засобами Protégé-OWL, було зроблено наступне:
– визначено область та масштаб онтології;
– проаналізовано можливість використання існуючих онтологій;
– описано важливі терміни в галузі матеріалознавства;
– визначено класи та їх властивості;
– внесено екземпляри класів.

Базовим елементом розробленої онтології, як і в більшості інших, є класи, які можна розглядати як множини, що містять об'єкти. Класи в цій онтології організовані в ієрархію клас-підклас, тобто таксономію. Наприклад, об'єкти «Чавун» і «Сталь» (рис. 1) відносяться до класу «Залізовуглецеві сплави», котрий в свою чергу, разом з «Металокерамічними сплавами» та «Сплавами на основі кольорових металів», є підкласом «Сплавів». У випадку побудови глибшої ієрархії об'єкти «Чавун» і «Сталь» можна розглядати як окремі класи зі своїми підкласами та об'єктами.

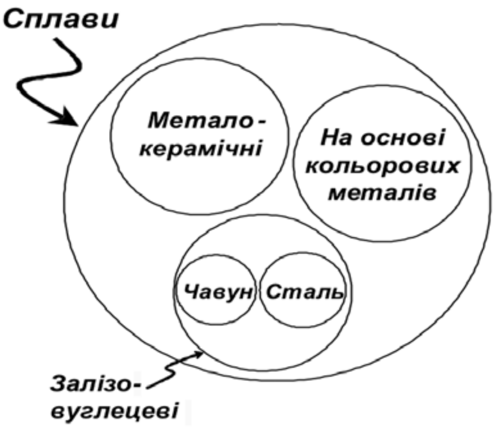


Рисунок 1 – Зображення структури класів

Після введення нового класу в онтологію потрібно задати його властивості, для того, щоб поняття ПрО були наповнені певним змістом і перебували у зв'язках одне з одним. Застосовані три типи властивостей, які подано на рис. 2. Так на рис. 2*а* властивість «*має домішку*» зв'язує об'єкт «сталь» класу «метали» з об'єктом «сірка» класу «неметали»; на рис. 2*б* – властивість «*має питому вагу*» зв'язує об'єкт «сталь» з числовим значенням «7800 кг/м^3»; а на рис. 2*в* як додаткова інформація

застосовується описова властивість «*використовується в*», котра зв'язує клас «сталь» з текстовими даними, наприклад, «будування літаків». Таким чином, при створенні нової властивості класу або об'єкта в Protégé-OWL задається відповідний її тип. Також властивості можуть бути оберненими, тобто мати інверсію. Наприклад, інверсія до властивості об'єкта «має колір» – «бути кольором».

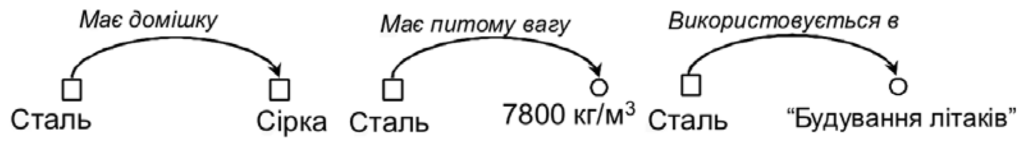


Рисунок 2 – Типи OWL властивостей в онтології в галузі матеріалознавства:
а – властивість об'єктів; *б* – властивість типів даних; *в* – анотована (описова) властивість

Після того, як введено основні класи онтології в галузі матеріалознавства, визначено їх властивості та встановлено умови їх взаємозв'язку, для забезпечення конкретизації ПрО вводяться екземпляри (об'єкти) класів. Об'єкти як найнижчий рівень онтології унаслідують всі властивості класів, до яких вони належать. Прикладами об'єктів в онтології матеріалів можуть бути конкретні марки сталей або назви чистих металів (наприклад: Ст3, сталь 20, 15Х2МФА, мідь, залізо). У випадку, якщо об'єкт має під собою інші об'єкти, він автоматично стає класом. Фрагмент вікна редактора Protégé-OWL з ієрархією класів у ПрО матеріалознавство подано на рис. 3.

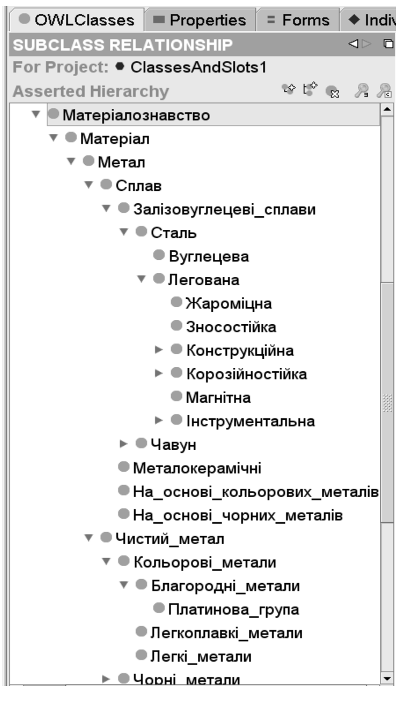


Рисунок 3 – Ієрархія класів онтології матеріалознавства

В розробленій онтології використовуються кванторні обмеження, числові, а також обмеження типу «має Значення». Кванторні обмеження побудовані з квантора (існування (∃) або універсальності (∀)), властивості та деякого класу об'єктів. Наприклад, вираз типу «∃ *Використовують* Конструкційні сталі» складається з квантора

існування (∃), властивості «*Використовують*» та класу об'єктів «Конструкційні сталі» і визначає набір об'єктів або клас, котрі зв'язані і можуть використовувати об'єкти з класу «Конструкційні сталі». А вираз типу «∀ має_легуючий_елемент Чистий_метал» визначає набір об'єктів або клас, всі елементи якого є зв'язані з елементами класу «Чистий метал». Цей вираз, застосований до класу «Леговані сталі», означає, що всі об'єкти цього класу мають легуючі елементи, котрі належать до класу «Чисті метали» (рис. 4).

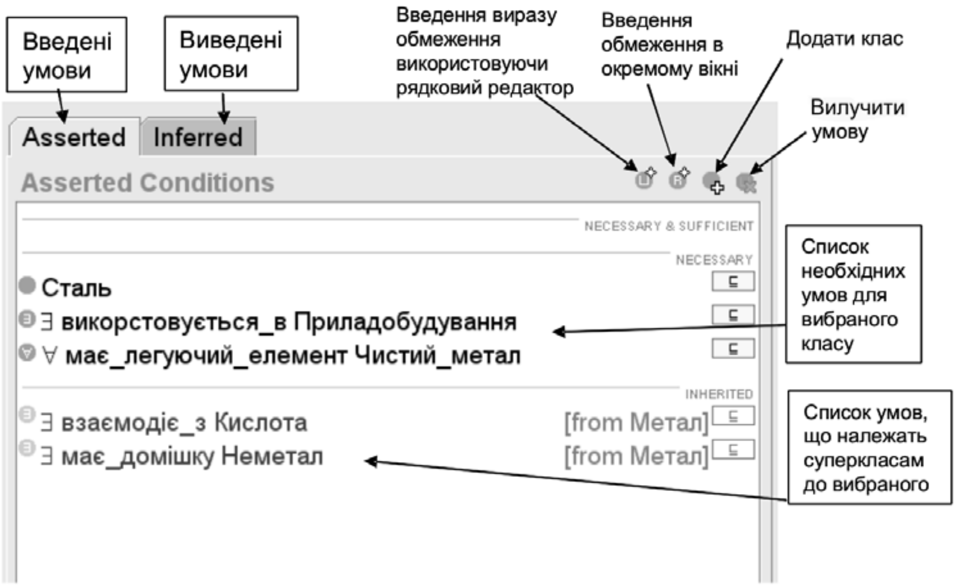


Рисунок 4 – Вікно введення виразу обмежень

Інший тип обмежень використаний в онтології – числовий клас об'єктів, що мають *не менш ніж*, *не більш ніж* або *чітко визначену кількість* зв'язків між іншими об'єктами або даними. Наприклад, за допомогою такої умови задано процентний вміст та максимальну кількість домішок у чистому металі, коли цей метал ще можна вважати чистим. *Обмеження типу «має значення»* позначені символом э і визначають клас або набір об'єктів, що, використовуючи деяку властивість, мають принаймні один зв'язок з конкретним об'єктом, тобто вираз «*має_Домішку* э Сірка» (де «Сірка» – об'єкт класу «Неметали») визначає набір об'єктів (наприклад, металів), котрі можуть мати домішкою саме сірку.

Описавши всі класи, властивості, обмеження і об'єкти ПрО, одержуємо базу знань, що є основою для побудови ІС, здатних здійснювати операції над інформацією. Створену онтологію в редакторі Protégé-OWL можна експортувати у формат, зрозумілий для інших редакторів, котрі використовуються безпосередньо для розробки інтелектуальних систем (CLIPS, HTML, RDF, OWL).

Практичне застосування онтології в галузі матеріалознавства у наукових дослідженнях

Розроблена онтологія є ядром інтелектуальної системи, котра забезпечуватиме пошук та семантичне розпізнавання інформації, що надходить з мережі Інтернет, нагромадження та класифікацію одержаної інформації та її використання в наукових дослідженнях.

З цією метою в роботі побудовано спеціалізовану метапошукову систему (МПС) матеріалознавчого профілю для автономного (без участі фахівців у галузі матеріалознавства) періодичного пошуку в мережі Інтернет наукових публікацій англійською мовою з цієї тематики. Система виділяє з онтології підмножину понять, що мають найбільшу вагу, і будує з них множину ключових слів для формування запиту до пошукових серверів Google, ScienceDirect за векторно-просторовою моделлю (рис. 5). Початково великий об'єм множини ключових слів ітераційно скорочується до отримання непорожньої множини посилань на текстові документи у відповіді пошукового сервера. З отриманої множини адрес виключаються повторення та адреси, документи за якими вже були отримані раніше. З результуючої підмножини адрес система отримує нові статті (і/або їх анотації), що потенційно можуть становити науковий чи комерційний інтерес для фахівців у галузі матеріалознавства – користувачів системи, у форматі текстових документів.

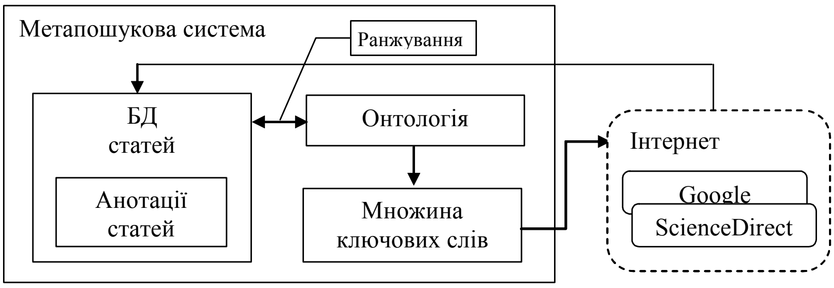


Рисунок 5 – Взаємодія метапошукової системи з веб-ресурсами

За допомогою побудованої онтології матеріалознавства отримані англомовні анотації наукових статей упорядковуються за їх релевантністю до ПрО. З цією метою система будує модель онтології авторів кожної зі статей, що розглядаються, використовуючи в якості прототипу власну онтологію, обчислює семантичну відстань між центром семантичної ваги власної онтології та онтології авторів кожної зі статей, вилучає з множини релевантних документи, для яких ця відстань більша за порогову та вносить усі інші до бази даних системи з визначеними таким чином коефіцієнтами релевантності [13].

Для побудови моделі онтології авторів статті МПС формує онтологію-прототип з нульовими ваговими коефіцієнтами понять та зв'язків між ними, вибирає з досліджуваної анотації ті поняття та зв'язки, які входять до онтології системи та послідовно присвоює їм у модельній онтології відповідні ненульові вагові коефіцієнти.

На наступному кроці процедури інформаційного пошуку до модельної онтології автора статті МПС додає множину понять, семантично тісно пов'язану з поняттями, вжитими даним автором та розпізнаними системою. Для цього в онтології МПС виділяється підмножина вжитих автором понять, а також порогове значення ваги семантичного зв'язку цих понять з іншими поняттями онтології МПС та вибирається множина понять контексту, зв'язок яких з поняттями, вжитими автором, не нижчий за пороговий. Дана множина понять контексту додається до модельної онтології автора публікації.

На заключному кроці розпізнаним поняттям та поняттям контексту в модельній онтології автора присвоюються вагові коефіцієнти з онтології МПС, а усі інші поняття та зв'язки з нульовими ваговими коефіцієнтами відкидаються. Модель

онтології автора статті може бути доповнена даними моделей попередніх статей даного автора шляхом суміщення відповідних графів.

Таким чином, сформована модель включає в себе інформацію як про зміст публікації, так і про наукові інтереси її авторів в системі координат наукових інтересів користувача метапошукової системи.

Серед інших можливих практичних застосувань онтології, розробленої в галузі матеріалознавства, виділимо:

1. *Використання онтології у процесі аналізу металографічних зображень.* Сьогодні в лабораторіях металу на підприємствах енергетики та нафтопереробної промисловості і відповідних відомчих науково-дослідних закладах нагромаджено велику кількість металографічної інформації, отриманої на різних етапах експлуатації конструктивних елементів різноманітного обладнання. Для підвищення ефективності її використання і подальшого аналізу виникає потреба перевести таку інформацію на електронні носії і зберігати її в комп'ютерній базі даних. Така база даних повинна постійно поповнюватись інформацією в процесі проведення планових обстежень. Використання бази даних металографічних зображень в складі інтелектуальної системи, ядром якої є онтологія матеріалів, дасть змогу використати з максимальною ефективністю одержані знання під час аналізу структурних змін в металі та виявлення пошкоджень [2].

2. *Визначення залишкового ресурсу.* Завданням ІС в цьому випадку є попередження про проблемний елемент конструкції, ймовірність його руйнування та обґрунтовування потреби позапланового обстеження. Це дасть змогу запобігати непередбаченим руйнуванням елементів конструкцій та упереджувати аварійні зупинки технологічного процесу.

3. *Використання онтології у процесі виготовлення нових конструкцій.* Оскільки в онтології зібрано найбільш повну інформацію про властивості матеріалу за різних умов експлуатації, його поведінку в різноманітних агресивних середовищах, інженер або конструктор, використовуючи базу знань, у процесі проектування може вибрати саме той матеріал, котрий найбільш буде підходити для вирішення конкретної задачі.

4. *Використання онтології під час проведення досліджень та планування експерименту*. У процесі планування експерименту ІС допоможе проаналізувати літературні дані з певної проблематики, оцінити результати досліджень аналогічних об'єктів та матеріалів, а також взяти до уваги досвід інших спеціалістів та науковців, які працюють у цій галузі.

Вище описано лише декілька практичних застосувань онтології матеріалів, однак цей перелік може бути значно ширшим та охоплювати майже всі галузі науки, де використовуються матеріали.

Висновки

Підсумовуючи, виділимо декілька головних напрямів застосування онтології у Фізико-механічному інституті НАН України:

– база даних, де відображені, наприклад, виробник металу/сплаву, його місцезнаходження, контактна адреса, ціна тощо;

– наочне відображення ПрО. Це зручно, зокрема, для молодих спеціалістів – швидко і повною мірою ознайомитись з ієрархією основних понять, термінів у галузі матеріалознавства;

– електронний архів для зберігання документів, класифікованих відповідно до УДК;

– ядро інтелектуальної системи.

Сьогодні наукові надбання інституту доцільно систематизувати в базу знань матеріалів, що буде зручним довідником для кожного користувача, а у майбутньому стане основою для створення першої загальнодоступної предметної онтології в галузі матеріалознавства.

Література

1. Gruber T.R. A translation approach to portable ontology specifications. Knowledge Acquisition, 1993. – Vol. 5. – P. 199-220.
2. Даревич Р.Р., Марков А.Д., Студент О.З. Створення експертної системи прогнозування пошкоджень в енергетиці та нафтохімії на основі металографічного аналізу // Машинознавство. – 2005. – № 2. – С. 48-53.
3. Sheng-Tun Li, Huang-Chih Hsieh, I-Wei Sun. An Ontology-based Knowledge Management System for the Metal Industry. WWW-2003 (Alternate Paper Tracks). – Budapest, Hungary, 2003.
4. Палагин А.В. Архитектура онтологоуправляемых компьютерных систем // Кибернетика и системный анализ. – 2006. – № 2. – С. 111-125.
5. Вдовіченко А.В. Інтелектуалізовані пошукові системи. Класифікація та порівняння // Искусственный интеллект. – 2002. – № 3. – С. 61-70.
6. The KAON. – Режим доступу: http://kaon.semanticweb.org/
7. The OilEd. – Режим доступу: http://oiled.man/
8. The OntoEdit. – Режим доступу: http://www.ontoprise.de/com/ontoedit.htm
9. The Ontosaurus. – Режим доступу: http://www.isi.edu/isd/ontosaurus.html
10. The OpenCyc. – Режим доступу: (http://www.opencyc.org/)
11. The Protege Project. – Режим доступу: http://protege.stanford.edu.
12. The Protege Project. – Режим доступу: (http://protege.stanford.edu/index.html)
13. Оцінка подібності текстових документів на основі визначення інформаційної ваги елементів бази знань / Р.Р. Даревич, Д.Г. Досин, В.В. Литвин, З.Т. Назарчук // Искусственный интеллект. –.2006. – № 3. – С. 500-509.

Д.Г. Досин, Р.Р. Даревич, Н.В. Шкутяк

Разработка онтологии материаловедения средствами Protégé-OWL

В статье на основе существующих подходов к построению онтологии нижнего уровня, а также форматов и стандартов подачи знаний представлена разработанная авторами технология построения онтологии в области материаловедения средствами Protégé-OWL. Онтология ориентирована на применение в интеллектуальных системах информационного поиска в текстовых массивах данных. Проанализированы возможности практического применения разработанной онтологии в научных исследованиях.

D. Dosyn, R. Darevych, N. Shkutyak

Development of Ontology in Materials Science by Protégé-OWL Tool

In the paper using the existing approaches to ontology construction of the lower level, and also formats and standards of knowledge representation, the technology of ontology developed in materials science domain by Protégé-OWL tool is presented. The developed ontology is directed to application in the intelligence systems of information search in the text data corpus. Possibilities of practical application of the developed ontology in scientific researches are analysed.

Стаття надійшла до редакції 17.07.2008.